Statistical Learning: Distribution-free Prediction and Confidence Intervals for Linear Regression Problems

Babolcsay, Barbara

Eötvös Loránd University
Faculty of Science

2024/01/08

Problem Setting

- Consider a linear regression system:

$$\left.\begin{aligned} \phi_1^T\theta^* + N_1 &= Y_1 \\ \phi_2^T\theta^* + N_2 &= Y_2 \\ &\dots \\ \phi_n^T\theta^* + N_n &= Y_n \end{aligned}\right\} \; \Phi^T\theta^* + N = Y$$

- The Least-Squares Estimate (LSE) of θ^*:

$$\hat{\theta}_n = (\Phi_n^T\Phi_n)^{-1}\Phi_n^T Y$$

- The question: how to build non-asymptotic distribution-free confidence regions around the point estimate?

Exact Confidence Regions

- $\hat{\theta}$ was the root of the normal equation:

$$0 = \sum_{t=1}^{n} \phi_t (Y_t - \phi_t^T \theta) = \sum_{t=1}^{n} \phi_t \phi_t^T (\theta^* - \theta) + \sum_{t=1}^{n} \phi_t N_t = H_0(\theta)$$

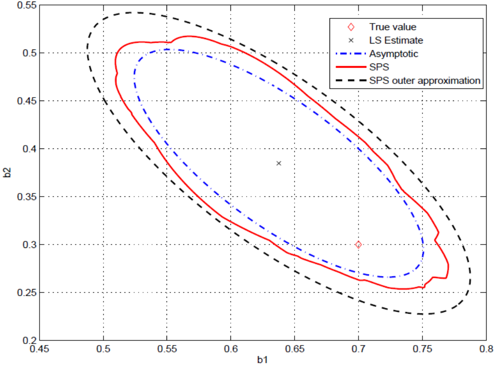
- Idea of SPS method [2]: perturb the signs of the prediction errors:

$$H_i(\theta) = \sum_{t=1}^{n} \alpha_{i,t} \phi_t \phi_t^T (\theta^* - \theta) + \sum_{t=1}^{n} \alpha_{i,t} \phi_t N_t$$

- If θ is "close" to θ^*, the original errors "size" is around the the perturbed ones - otherwise, it dominates among them (if we make $m-1$ perturbations and the error size is at most the $(m-q)th$ smallest then $\theta \in \hat{\Theta}$ wich is the confidence region with $p = 1 - \frac{m}{q}$).

Outer Approximation

- To get regions that are easier to calculate: ellipsoidal outer approximation
- It leads to $m - 1$ convex minimization problem and the *qth* largest optimum gives the proper radius of the ellipsoid.



[2]

From the Parameter Space to the Function Space

- What we have: confidence ellipsoids in the parameter space ($\mathbb{R}^d$), what we want: confidence bands in the function space ($\mathbb{R}$)
- This is a convex max/min problem on an ellipsoid that has an analytic solution which can be found e.g. using Lagrangian relaxation

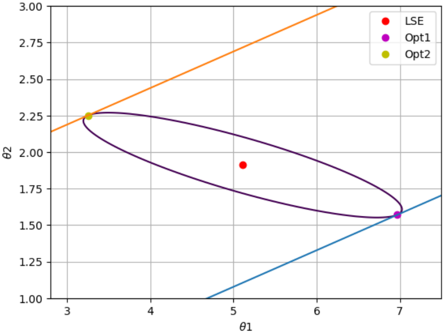


Figure: Example of a max/min problem in the parameter space, d=2

Results in the Function Space

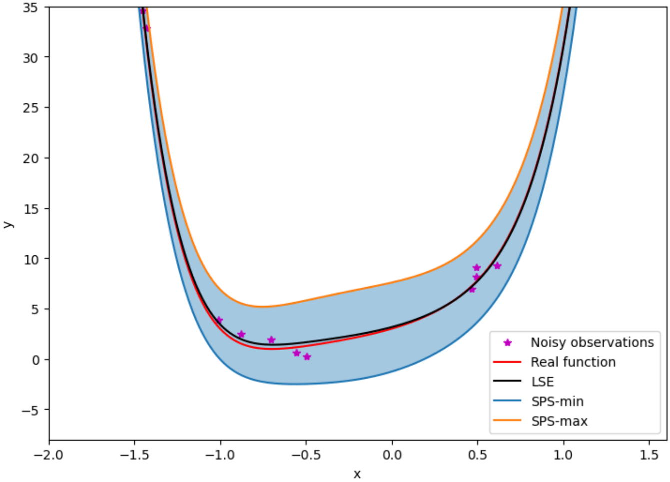


Figure: Polynomial regressors, n=30, d=7, p=0.8

Results in the Function Space

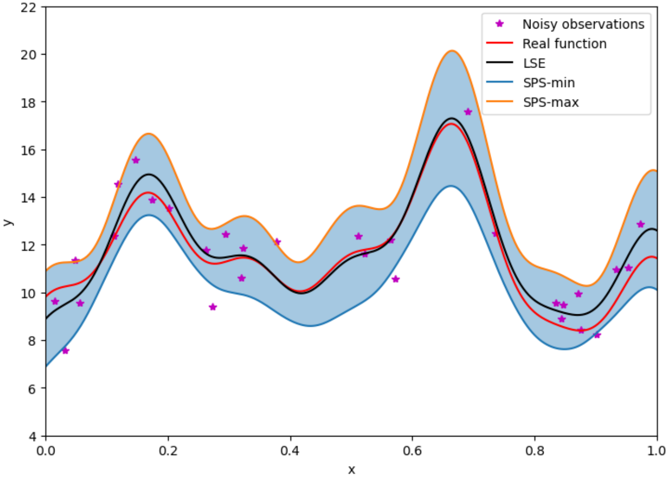


Figure: Gaussian regressors, n=30, d=7, p=0.8

What else can we use SPS method for?

- Block SPS [2]
- Modify to Regularized Linear Regression [1]
- Define the uncertainty for models given by kernel methods[3]

References

B. C. Csáji.
Non-asymptotic confidence regions for regularized linear regression estimates.
In *Progress in industrial mathematics at ECMI 2018*, volume 30 of *Math. Ind.*, pages 605–611. Springer, Cham, [2019] ©2019.

B. C. Csáji, M. C. Campi, and E. Weyer.
Sign-perturbed sums: a new system identification approach for constructing exact non-asymptotic confidence regions in linear regression models.
IEEE Trans. Signal Process., 63(1):169–181, 2015.

B. C. Csáji and K. B. Kis.
Distribution-free uncertainty quantification for kernel methods by gradient perturbations.
Mach. Learn., 108(8-9):1677–1699, 2019.

Thank You for Your Attention!